红色人物资源数字叙事框架及其实证研究*

徐孝娟 [1] 孙爱华 [1] 陈涛 [2]

（1 安徽大学管理学院 合肥 230039；2 中山大学信息管理学院，广州，510006）

摘要：[目的/意义]数字叙事作为一种在线叙事，有利于增强红色人物资源的可解释性与人机互动性，为优化红色人物资源知识化开发提供可能 **[方法/过程]**在梳理历史人物资源知识化开发与红色人物相关研究基础上，提出由叙事资源整合层、叙事元素抽取层和叙事作品生成层组成的红色人物资源数字叙事基本框架，紧接着以“皖西红色人物”为例，将知识图谱作为叙事工具进行实证研究。**[结果/结论]**红色人物资源数字叙事框架可完善目前主要依据红色资源“内容”和“外部”两个层次设计实体语义关联的局限，形成了可操作性的结构流程，为红色资源知识化开发提供借鉴；较传统人工抽取与规则匹配的知识抽取方式，基于本体与BERT 模型的三元组批量抽取模式提高了红色人物资源知识抽取效率；借助知识图谱实现了红色人物资源可视化和主题叙事的知识发现，佐证了数字叙事在红色人物资源知识开发和组织中的应用价值。

关键词：红色人物资源 数字叙事框架 知识图谱 主题知识发现

分类号：G350

1 引言

习近平总书记多次对红色文献资源开发利用提出要求，并上升至国家整体文化战略层面，强调要用好红色资源，赓续红色血脉，努力创造无愧于历史和人民的新业绩[1]。近年来，红色文献资源整合利用和开发已不在局限于元数据等外部特征的浅层次描述，而是深入到资源内部利用实体与关系构成“知识元”进行表征，进而通过挖掘资源中“知识元”的语义关联以实现增值[2]。红色人物作为红色文化的缔造者和传播者，是红色资源内容呈现和展演的主要载体[3]，林语堂言“史源于事，事源于人，无人则无事，无事则无史”。然而，囿于数字技术应用水平不高和多模态特征识别不全的现实难题[4]，传统平面与线性叙事已无法满足当前用户的多元需求，为此红色人物资源的传统叙事亟待向数字叙事转化，特别是历史语境下红色人物资源叙事的可解释性和资源平台的人机交互性等方面仍有待完善。

叙事理论作为研究“在时间和因果关系上意义有着联系的一系列事件的符号再现”相关现象的理论总称，通过其空间-道德-互动-跨媒介等叙事体系为知识组织提供理论指导，可生动地再现革命文化史料之间的关系，完整呈现历史史实，避免支离破碎式解读，使红色文献资源价值的发掘更加充分[5-7]。为此，本研究在考虑红色人物资源来源、类型、内容及结构等多样性特征基础上，引入叙事理论探索资源整合、元素和元素关系抽取以及知识应用的红

*本文系安徽省教育厅重点项目（项目编号：2022AH050022）、2022 年安徽省高校优秀青年人才支持计划项目（项目编号：gxyqZD2022002）的研究成果之一。

作者简介：徐孝娟，副教授，博士；孙爱华，硕士研究生；陈涛，副教授，博士。

色人物资源数字叙事框架；并以“皖西红色人物数据”为例，探索红色人物资源知识关联，重构人-事-地-时等实体场景，完成红色人物资源的知识聚合，形成融合叙事与资源内外部特征的多层次知识组织与服务体系。

2 文献回顾

2.1 历史人物资源知识化开发相关研究

数智技术发展推动了历史人物资源知识组织与智慧服务的进程，其知识化开发的相关理论与实践不断发展，其研究成果逐渐由粗粒度的文献整理到细粒度的知识组织转变，历史人物资源组织与数字人文研究的知识化开发程度较大提高。目前国内外研究主要聚焦于本体研究、知识图谱构建、知识库建设三个维度。

第一，本体研究方面，主要关注如何构建历史人物本体模型，如先秦人物知识本体模型和靖江王府人物关系等，并应用构建的历史人物本体模型进行知识抽取、人物关系推理，进而改善用户查找知识的体验[8-10]；第二，知识图谱方面，如学者分别聚焦于历史人物多层关系网络来识别历史人物间的不同派系[11]、构建人物社会网络图谱对宋代浙江中部地区梧州的亲属与学社变迁[12]、历史名人游学足迹关系数据模型[13]、人物履历知识图谱的自动化构建方法[14]及数字人文分析工具探索名人日记资源挖掘[15]，并开始注意隐藏关系的推理[16]，该维度研究注重实证检验，其案列较为多元；第三，知识库建设方面，学者多探索有关人物专题库的建设思路和优化途径[17]，如利用关联数据技术对图书馆特藏资源进行开发与重组，构建了家谱知识服务平台，该平台不仅支持寻根搜索服务，还为科研用户提供数据分析支持[18,19]，以及学术名人知识库建设的基本模式，并以吴宝康学术名人知识库为例探讨实践探索中的技术与操作流程[20]。

2.2 红色人物资源相关研究

近年来，红色人物资源整合开发与研究工作受到日益重视，图档博等文化记忆机构根据当地或本馆特藏陆续开展红色人物资源的整理并提供在线服务，代表性平台如陕西省图书馆的陕甘宁边区红色记忆·人物库[21]、复旦大学图书馆的陈毅专题库[22]、上图红色文献平台的中共组织史专题库[23]等。不难发现，红色人物资源正在发生“藏”到“用”的转变。与此同时，现有研究多关注红色人物资源库的建设及其资源组织与应用，具体为：第一，红色人物资源库的建设。窦鹏以陕甘宁边区红色记忆人物库为例，介绍了红色人物数据库构建的流程与经验，为红色人物数据库构建提供参考和借鉴[21]；孙烈涛以周恩来图书馆数字化建设为例，提出了红色教育资源数字化服务建设模式[24]。第二，资源组织与应用。张云中等提出了基于知识谱图红色历史人物知识问答服务框架，并以老上大历史人物资源为语料库进行了平台搭建，实现了红色历史人物资源的细粒化组织与知识服务[3]；刘伟丽构建中共一大人物知识图谱，实现了人物、事件、地点、机构等要素的语义关联，为红色人物资源的语义化组织提供参考[25]。

综上所述，红色人物资源的在资料整理、数字化以及专题库建设已经取得了初步成果，但仍存在较大探索空间：首先研究视角大多从红色文献资源开发的整体考虑，红色人物作为红色文化的重要元素或主题，对其深入挖掘有待扩展；其次，红色人物资源知识表达体系的构建中多将资源置于数据要素层面，难以构建复杂的知识体系，以及难以解读传达革命文化蕴含的深刻内涵，对于红色文献主题聚类、语义关联与知识发现等沉浸式挖掘有待加强；此外，部分学者逐渐发现数字叙事对红色文献资源研究的重要性，并探索了红色档案资源的叙事表达、叙事性开发模型及多模态红色馆藏开发的优化策略[4,7,26,27]，然而对于红色人物资源叙事逻辑及一般性框架的探索，特别是实证研究尚有待加强。基于此，本文引入叙事理论，契合资源开发需求，设计红色人物资源数字叙事框架，并以皖西红色人物为例进行实证研究，为多模态红色人物资源知识化、场景化开发提供借鉴。

3 红色人物资源数字叙事框架设计

本文在参考 Barber J F、付雅明、张斌等学者[28-30]对叙事理论探讨、框架搭建和研究方法基础上，结合红色人物资源开发需求与用户体验，归纳出红色人物资源数字叙事的必备要素、实施过程与应用场景；与此同时，借助叙事理论协同完善红色人物资源知识本体核心词表，完善目前主要依据资源“内容”和“外部”两个层次设计实体语义关联的局限。为此，形成面向数字叙事的红色人物资源知识化开发理论框架，该框架分为叙事资源整合层、叙事元素抽取层与叙事作品生成层，为后续实证研究提供理论指导，具体见图 1。

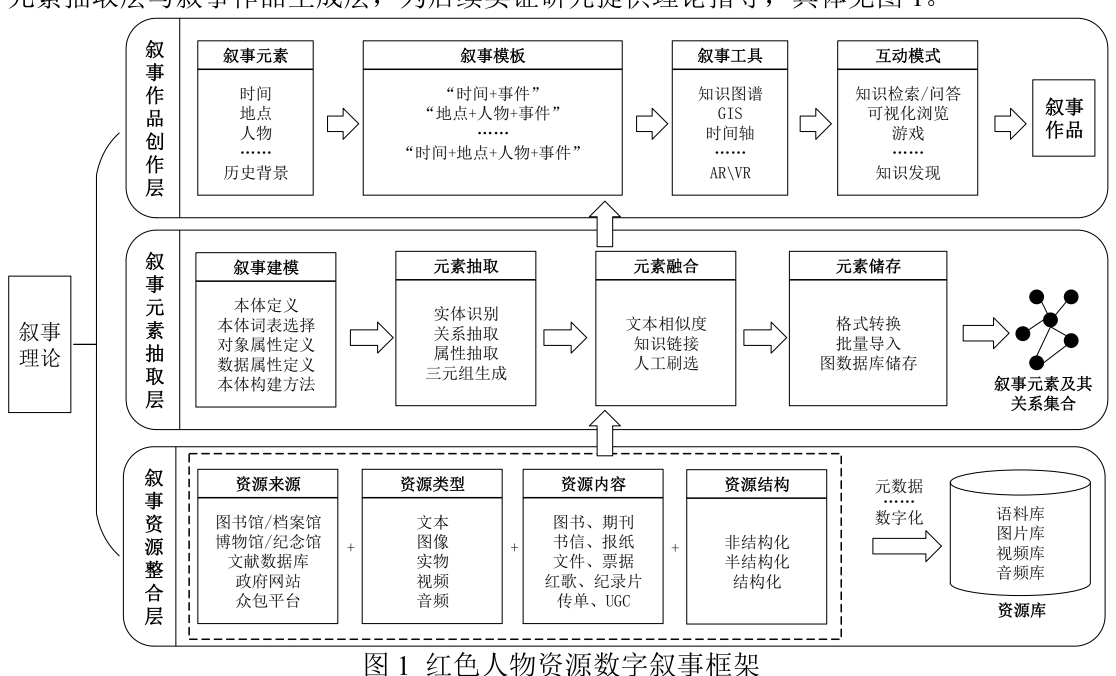


图 1 红色人物资源数字叙事框架

3.1 叙事资源整合层

叙事资源整合层主要目标是完成对不同来源、类型、内容及结构等红色人物资源的整合，使其数字化、文本化与资源化，为数字叙事提供数据基础。具体步骤如下：红色人物数据因历史原因，其资源来源较为分散，资源类型、内容多样，据此按照数据结构化程度，将红色

人物资源分为非结构化数据、半结构化数据与结构化数据；并在此基础上完成红色人物数据的预处理与资源库构建。值得注意的是：（1）原始红色文献属于非结构化数据，则经过字符识别或人工标注等方式才能实现文献的数字化和文本化；（2）进一步的整合知识百科与官方网站等相关红色人物词条记录，此类数据具有价值高、分布广和多模态的特点；（3）红色人物专题库已进行高度数字化整合，其数据较为规整与结构化，数据中冗余的部分较少，可信度较高，可直接作为数据分析的材料。

3.2 叙事元素抽取层

叙事元素抽取层的主要功能是借助本体与知识抽取相关技术实现红色人物资源最小叙事单元的处理与生成。在整合叙事资源的基础上，参考知识抽取流程[31-33]，将叙事元素抽取流程划分为叙事建模、元素抽取、元素融合与元素储存，为此完善资源"内容"和"外部"两个层次设计实体语义关联的局限。具体如下：

（1）元素建模完成红色人物数字叙事概念层的设计，即红色人物资源本体构建，是红色人物数字叙事的基本模板，通过对红色人物资源相关术语及其属性进行梳理与定义，实现领域知识的标准化与规范性。（2）以红色人物本体模型为依据对红色人物资源进行叙事元素抽取，可分为实体抽取、属性抽取与关系抽取。其中实体抽取是指从文本中抽取现实存在并具有相互区别的一类物体，例如人物、时间、地点与事件等；属性抽取通常对某一实体的相关属性进行补全，以事件实体为例，补充事件的发生地、发生时间或相关人物等；关系抽取通常会提取不同实体之间的关联关系，且这些关系具有一定的方向性。（3）考虑到红色人物数据的多源异构性，知识表征的形式也呈现出多样性，因此，需要对抽取的元素与关系实施分析、整合和消歧等知识融合。例如：共指问题，多个不同描述指代同一叙事元素，例如同一个人物的不同称呼或由于历史变迁所产生的地名变更；歧义问题，叙事元素名称可以指代多个真实实体，导致表意不清。（4）元素储存，首先将抽取的叙事元素、属性与关系转换成 SPO 三元组形式，如<元素，关系，元素>与<元素，属性，属性值>，构成描述红色人物资源最小知识元；其次将生成的数据批量导入到数据库进行储存。

3.3 叙事作品创作层

叙事作品创作层主要功能在于利用以图结构为主的叙事模板，借助叙事工具与普通用户、领域专家进行互动，最终实现红色人物资源叙事作品的创作并加以呈现，其核心要素包括叙事元素、叙事模板、叙事工具以及互动模式[4,30,34,35]。（1）叙事元素作为构成叙事作品的基本单元，包括但不限于时间、空间、人物、事件、物体等，其中不同红色人物资源的元素之间按照某种规则组合可呈现出不同的红色人物相关故事。（2）叙事模板也称叙事结构，是由一系列特定背景、人物、事件等红色人物资源叙事要素组成的较为稳定的抽象结构，引导用户从一个节点到另一个节点，实现不同红色人物故事素材的关联与组合。（3）叙事工具作为用户与资源互动的媒介，可有效渲染并呈现叙事作品，可根据叙事结构与情节不同，选择合适的叙事工具。例如，协助人文研究的知识图谱、GIS 与时间轴，以及提供娱乐的游戏

软件、AR/VR 等；（4）互动模式是用户使用叙事工具的不同方式，相较于传统叙事，数字叙事具备更强的交互性，主张将叙事过程的主导权让渡给用户，不同用户根据自身需求进行知识检索、问答、可视化浏览、游戏以及知识发现等。

4 皖西红色人物资源整合与元素抽取

4.1 叙事资源获取与工具选择

4.1.1 数据简介与获取

本文选取“皖西红色人物”资源作为案例进行实证研究，其具有可行性、可操作性及代表性：其一，皖西地区是江淮分水岭，具有“襟江带淮，承东接西”重要的地理位置，产生了丰富的红色文化和革命传统，被誉为红军、将军和共和国的“三大摇篮”，皖西红色人物数据具有重要的史料研究价值[36]；其二，数字资源中的红色人物所属时间跨度较长、地点分布较广，包括了多时期多地点的红色英雄，且人物经历描述非常完整，数据量较为丰富；其三，数字资源经过权威部门编撰与审核，具有权威性和准确性。与此同时，由于数据总量较大，笔者使用 python 的 scrapy 爬取并解析上述资源中红色人物的多模态数据，涉及文本、图片、音视频的链接地址等，最终共获取 676 条红色人物数据，并储存为 JSON 文件共后续使用。

4.1.2 叙事工具选择

本文选择知识图谱作为叙事工具。知识图谱，一方面具备成熟的理论与技术体系，能够为红色人物资源提供的数字资源知识聚合范式参鉴；另一方面，作为一种图形化、关联化的知识表现形式，不仅能够实现红色人物资源的网状叙事，而且其强大的知识检索与可视化功能可以为人机交互提供条件，有益于实现红色人物资源的历史场景再现与知识发现。

4.2 叙事建模

当前研究多从资源“内容”和“外部特征”两个层次设计实体语义关联，可以较好的描述资源的基本内容但整体叙事性较弱。数字叙事的引入使得红色人物资源从单向、线性叙事向网状叙事转变，促进资源的叙事性、细粒度与关联性开发，为叙事建模的整体设计与标准化组织提供理论指导。本文将实用性较强的七步法进行红色人物资源本体构建，具体步骤如下：

（1）确定本体适用范围和查找可复用本体。第一，本体主要用于红色人物资源的知识组织，以红色人物作为数据源，描述红色人物基本信息及其生平履历。第二，复用本体是指重复使用现已公开成熟本体中类或属性的过程，有助于减少概念重复定义，降低本体构建的复杂度与成本，增强本体可扩展性。本文通过对关联开放词表（Linked Open Vocabularies ,LOV）进行检索得到 33 个与人物相关的本体词表，并根据各人物本体的特征筛选出相关本体 5 个，包括 rel（描述人与人之间关系的本体）、bio（描述人物传记信息的本体）、foaf（描述互联网社交信息的本体）、whois（描述个人资料或履历的本体）、shoah

（描述大屠杀相关的概念的本体）。其中可复用的类有："Person"、"Organization"、"TemporalEntity"、"Place"、"Event"等，可复用的属性有："birth"、"death"、"gender"、"otherName"、"event"等。

（2）概括红色人物资源本体核心概念和分类体系。数字叙事的核心目的在于通过不同的叙事逻辑，借助数字技术最大程度、最细粒度的还原历史全景，因而需要总结并提炼多特征与更细知识粒度的概念及分类体系，具体如下。第一，根据红色人物资源的文本特征，其核心概念：① 红色人物，红色人物资源的叙事主体；② 红色人物的人生履历。人物的履历一般包含求学经历与战争经历，与之相关的概念有战争事件、重大会议、红色地点，其中可抽取出人物、组织机构、时间、事件、地点、资源等元素。第二，采用自顶向下方式进行构建，对已完成的顶层概念再进行类别划分，确定概念之间的父子关系，以形成红色文献本体的概念分类体系。顶层概念实体包括"人物"、"组织机构"、"事件"、"地点"、"资源"、"时间"，其中"组织机构"按照其成立的性质被分为学校、军事机构、党组织和政府机构；"事件"被分为战争事件、重大会议、爱国运动以及重要组织机构成立；"地点"实体按照国家行政区划进行分类；"资源"按照其载体类型分为实物资源与数字资源。分类体系见表 1。

表 1 红色文献本体概念分类体系

一级类目	标签	二级类目	标签
Person	人物	-	-
Organization	组织机构	School	学校
		Party	党组织
		Army	军事机构
		Government	政府机构
		CulturalOrganization	文化机构
Event	事件	War	战争事件
		Conference	重大会议
		Campaign	爱国运动
		Establishment	重要机构的成立
Place	地点	-	-
Resource	资源	PhysicalResource	实物资源
		DigitalResource	数字资源
OfficialPosition	官职	-	-
TemporalEntity	时间	Instant	时间点
		Inerval	时间段

（3）定义红色人物资源的语义描述，即属性关系定义。属性关系定义需参考红色人物资源的虚拟场景，剖析人物生活的特定环境与影响事件发展的关键节点及其关联。本体属性

关系由数据属性与对象属性组成。第一，数据属性。数据属性如表 2 所示，包括：①人物描述，有姓氏、名字、字、性别、生年、卒年、受教育程度等，可表达较为完善的人物基本信息与背景，为红色人物叙事做铺垫；②资源元数据，描述资源的外部特征；③时间与地点，其中时间属性将“年月日”进行拆分与组合，地点属性则按照国家地理的行政区划（国家-省-市-县/区）进行划分，可适应叙事中较为复杂时空表达。第二，对象属性。对象属性规定了类与类之间的关系，如表 3 所示，包括①与人物相关，有出生/死亡日期是、籍贯是、学习于、参加、加入，有利于描述红色人物的人生履历，塑造完整的人物形象，并融合到叙事中；②与资源相关，有参考于、馆藏于、馆藏地是；③与事件相关，包括发生于（时间）与位于（地点）。这些语义关系有利于促进各关键节点在叙事系统中的相互联结，实现“人-事-时-地-资源”等实体之间的微观叙事。

表 2 数据属性

数据属性	描述	域	范围
lastName	姓氏	Person	xsd:string
firstName	名字	Person	xsd:string
courtesyName	字	Person	xsd:string
Gender	性别	Person	xsd:string
Education	教育	Person	xsd:string
imageUrl	图像链接	Person	xsd:angURL
Identifier	标识符	Resource	xsd:string
Title	标题	Resource	xsd:string
Content	内容	Resource	xsd:angURL
Description	资源描述	Resource	xsd:string
Format	资源格式	Resource	xsd:string
Carrier	资源载体	Resource	xsd:string
Date	日期（YYYY-MM-DD）	Instant	xsd:datetime
yearMonth	年-月（YYYY-MM）	Instant	xsd:gYearMonth
monthDay	月-日（MM-DD）	Instant	xsd:gMonthDay
Year	年（YYYY）	Instant	xsd:gYear
Month	月（MM）	Instant	xsd:gMonth
Day	日（DD）	Instant	xsd:gDay
Country	国家	Place	xsd:string
Province	省	Place	xsd:string
City	市	Place	xsd:string
district/county	区/县	Place	xsd:string

表 3 对象属性

对象属性	描述	域	范围
Birthday	出生日期是	Person	TemporalEntity
Deathday	死亡日期是	Person	TemporalEntity
nativePlace	籍贯是	Person	Place
Join	加入	Person	Organization
takePartIn	参与	Person	Event
referFrom	参考于	Person	Resource
Comrade	同志是	Person	Person
Relative	亲属是	Person	Person
Undertake	担任	Person	OfficialPosition
heldBy	馆藏于	Resouce	Organization
heldIn	馆藏地是	Resouce	Place
Occur	发生于	Event	TemporalEntity
locatedIn	位于	Event	Place

（4）构建红色人物资源本体。基于上述核心词表及其属性关系，利用本体构建工具 protégé进行本体构建，并以 RDF/XML 的格式进行储存，本体的结构如图 2 所示。

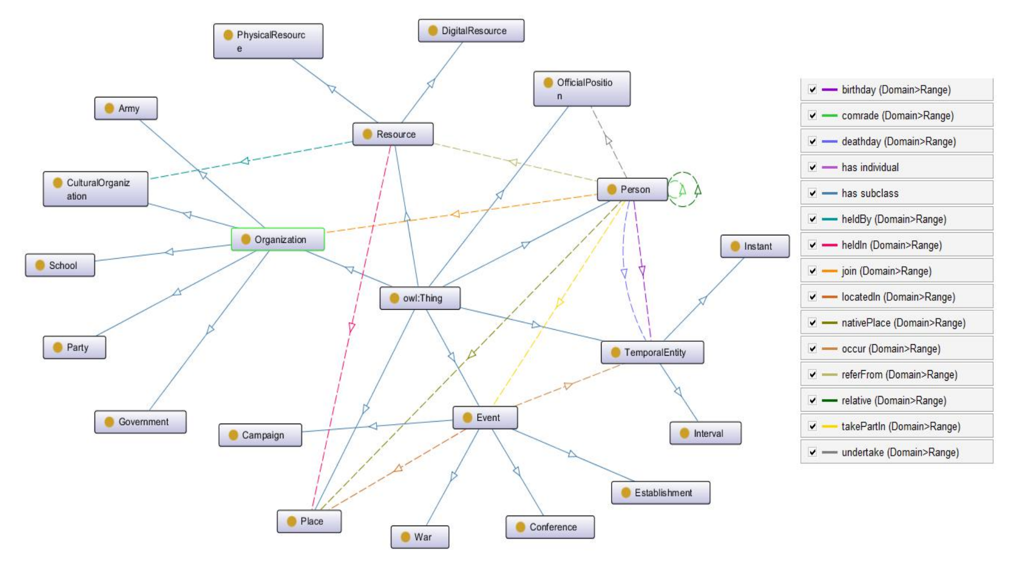


图 2 红色人物资源本体

4.3 元素抽取

4.3.1 算法模型结构

第一，考虑到红色人物资源中存在大量以红色人物为 Subject 的三元组结构，本研究突破已有“在命名实体识别的基础上进行关系分类”的常见关系抽取方式，参考语言模型概率图框架，从而更好地处理“同一（Subject，Object）对应多个 Predicate”的问题，优化红色人物资源的三元组抽取。其概率模型如公式 1（y1~yn 表示预测的输出序列）：

$$P(y_1, y_2, \dots, y_n | x) = P(y_1|x)P(y_2|x, y_1)\dots P(y_n|x, y_1, y_2, \dots, y_{n-1}) \quad (1)$$

具体思路：首先通过 Encoder 的输出 x 加上序列的起始标记来预测第一个单词，然后假设第一个单词已知来预测第二个单词，依此递推，直到出现结束标记，其模型如公式 2 所示：

$$P(s, p, o) = P(s)P(o|s)P(p|s, o) \quad (2)$$

第二，为了进一步处理“多对多”（多个 Subject、Object、Predicate）问题，本研究需要采用“半指针-半标注”结构进行再处理。为此，采取 BERT-半指针半标注的方式直接进行三元组抽取，主要逻辑：在 BERT 预训练模型上直接进行 fine-tune，即先采用预训练模型对所输文本进行 embedding，然后使用“半指针-半标注”的方式预测对应的 Subject 的首、尾位置；之后将所预测的 Subject 对应的 embedding 与 BERT 输出的 hidden-states 进行拼接，继续预测 Predicate 和 Object。值得注意是，在实际预测过程中可能存在多种 Predicate，因此对于每一种 Predicate，分别构建“半指针-半标注”的结构来预测对应的 Object 的首、尾位置，模型直接输出结果为三元组形式即(S,P,O)。其中，P 是给定的，根据标注的内容进行预测。模型结构如图 3。

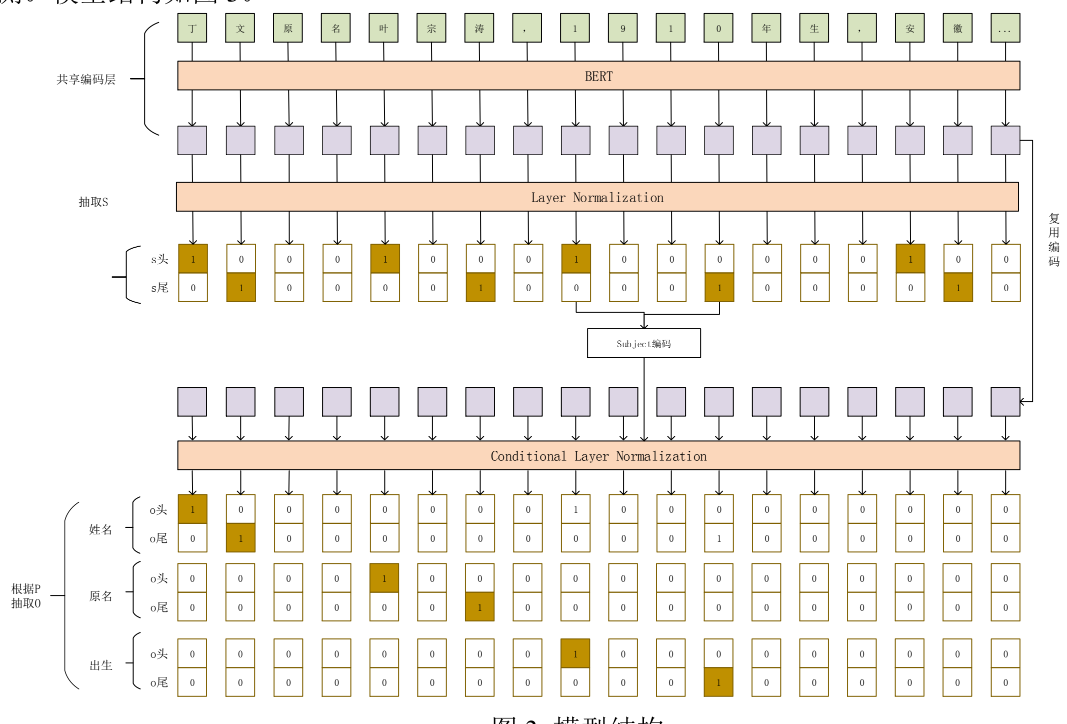


图 3 模型结构

4.3.2 文本标注与模型训练

首先，确定抽取的实体与关系。根据当前数据源的特征与红色人物本体的类与属性来确定需要人工标注的实体、属性与关系，其中人物、出生年份、死亡年份、其他名称、性别、籍贯、机构、事件、职位等实体分别用标签{Per,Bir,Dea,Oth,Gen,Loc,Org,Eve,Ran}进行表示，而出生于（年份）、逝世于（年份）、又名、性别是、籍贯是、学习于（机构）、加入（机构）、参加（事件）、担任（职务）、同志是（人物）等属性或关系分别用标签{birthyear,

deathyear,othername,gender,nativeSpace,study,join,takePartIn,undertake,comrade}进行表示。其次，进行数据标注。训练数据采用开源的皖西红色人物履历资源，并使用 Doccano 开源预料标注工具对分割好的人物数据进行标记，标记完成后，共获取到红色人物数据 1000 余条，将标记的样本以 8：1：1 的比例划分训练集、测试集与验证集。最后，将训练集与测试集导入到模型中进行训练。模型的超参数如表 4 所示。

表 4 模型超参数

超参数	描述	参数值
maxlength	输入句子的最大长度	512
batchsize	每轮实验的样本批量	32
epoch	最大迭代次数	20
Learning_rate	学习率	1e-4

4.3.3 模型评价

本实验采用准确率(Precision)、召回率(Recall)与 F1 值对 BERT 模型及其变体 ALBERT 与 RoBERTa 进行评价，结果见图 4，Precision、Recall 及 F1 值分别为 0.832、0.801 及 0.814。三个模型中 BERT 模型相对性能较差，而 ALBERT 与 RoBERTa 性能较为接近，其中 RoBERTa 模型效果最好，这说明该模型能够较为准确地实现红色人物资源三元组抽取。因此，本文选择 RoBERTa+“半指针-半标注”对皖西红色文献资源进行批量三元组预测并储存。

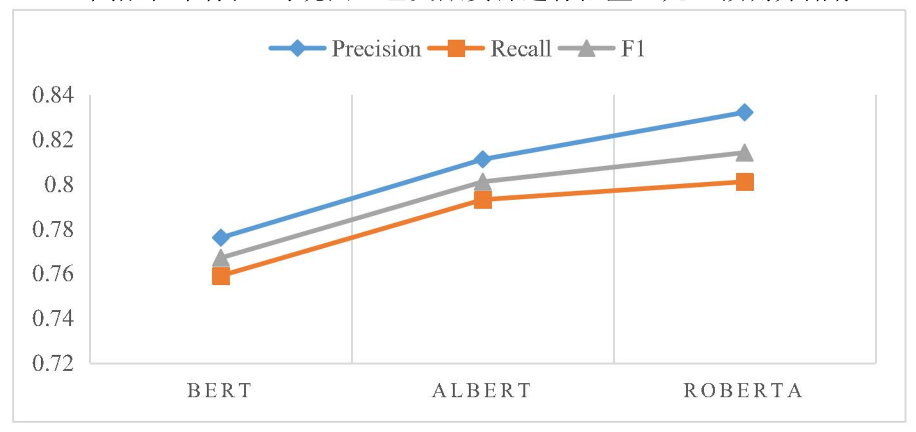


图 4 红色人物文本实体关系抽取模型评价

4.4 元素融合

实体融合方法主要包括文本相似度计算[3]、基于标准化词典或第三方数据库的知识链接[37]以及人工筛选与替换[38]，本文借鉴上述研究的处理方式，采用混合方法进行实体融合，具体方案如下：首先进行数据预处理，剔除实体名称中空格以及特殊符号并使用正式命名代替缩写与昵称。其次，根据实体类型采用合适的知识融合方法。①地点实体融合，地点实体中存在同一地点表述存在差异或地区行政级别的表达的层级不同，如“金寨”、“金寨县”、“安徽省金寨县”。为了解决这一问题，本文选用中国省市县区域公开数据集作为规范词表，再将抽取的地点实体与规范词表进行知识链接。②人物实体及机构实体融合，上海图书馆开

发的中共组织史数据库是目前较为完整且规范的组织史资料，本文将其作为人物与机构实体的知识库进行知识链接。③剩余实体本文采用 hanLP 工具包提供的短文本语义相似度模型来计算两个字符串的相似度，该模型经过大规模语料库训练，具有较高的准确率。

以“六霍起义”与“六霍农民起义”为例，由图 5 可知模型给出的相似度约为 0.967，有研究认为两个字符串相似度大于 0.85 时[3]，可认为其属于同一实体。最后，对实体融合结果进行人工评估并校正。



图 5 字符串相似度程序及示例

4.5 知识图谱生成

（1）经过上述相关步骤，本研究共获得实体 3247 个、属性 2905 条、关系 4035 条，具体描述见表 5、6 所示。

表 5 实体、属性描述及其数量

实体名	属性名	属性描述	数量
人物	姓名	人物姓名	676
	别名	人物的字、别名、化名等	162
	出生时间	人物出生年份	514
	逝世时间	人物逝世年份	515
	性别	人物性别	676
	籍贯	人物的祖居地或原籍	362
事件	-	--	399
学校	-	-	150
职位	-	-	1717
组织机构	-	-	305

表 6 语义关系描述及其数量

关系名	语义描述	头实体	尾实体	数量
加入	某人曾加入某个组织机构	人物	组织	823

参与	某人曾参与某个事件	人物	时间	815
学习于	某人曾学习与某个学校	人物	学校	221
担任	某人曾担任过某个职位	人物	职位	1927
同志是	某人是另一人的同志	人物	人物	249

（2）三元组数据存储。本研究选择目前主流的图数据库 Neo4j，并借助 py2neo 第三方库，将经过知识融合的三元组数据以“实体，属性，属性值”与“实体，关系，实体”两种形式批量导入到 Neo4j 中，最终生成包含 676 个皖西红色人物信息的知识图谱，部分知识图谱见图 6。

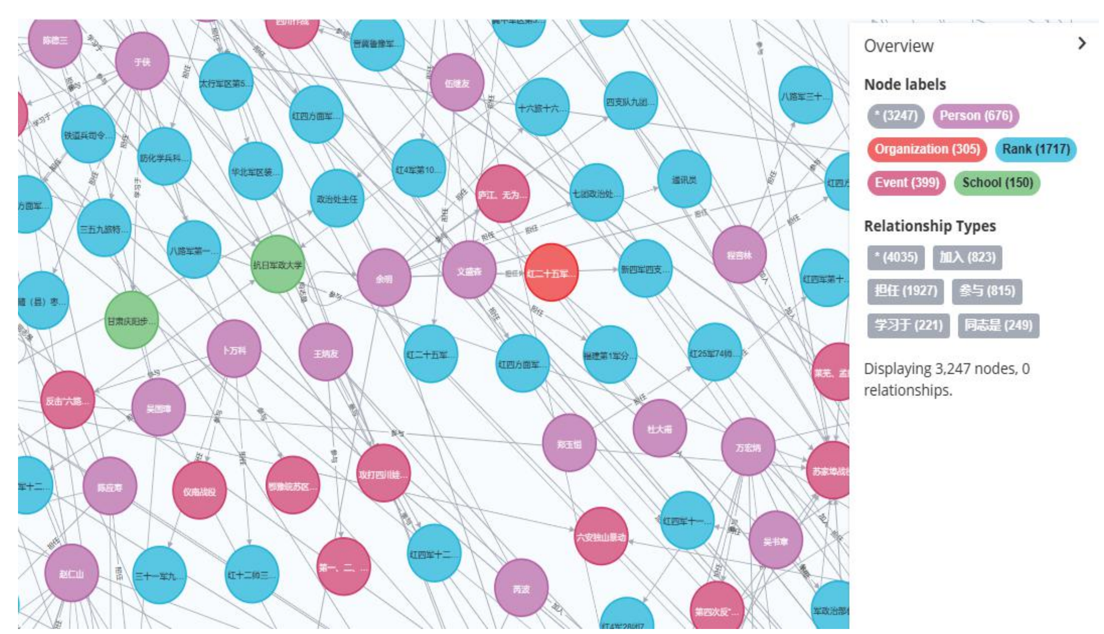


图 6 皖西红色人物知识图谱（部分）

5 皖西红色人物知识图谱数字叙事应用初探

5.1 知识图谱检索与可视化

检索与可视化功能作为知识图谱基本应用，Neo4j 图数据库集成了用于检索与可视化的图形交互界面，用户可以通过拖动、放缩与点击等操作直接获取图谱中的知识，并可基于 Cypher 语言构建检索式获取所需图谱。以洪学智将军的为例，查询图数据库中的相关节点与关系构成履历图谱，其查询检索式为“match p=(n:Person{name:'洪学智'})-[]->() return p”，结果见图 7。

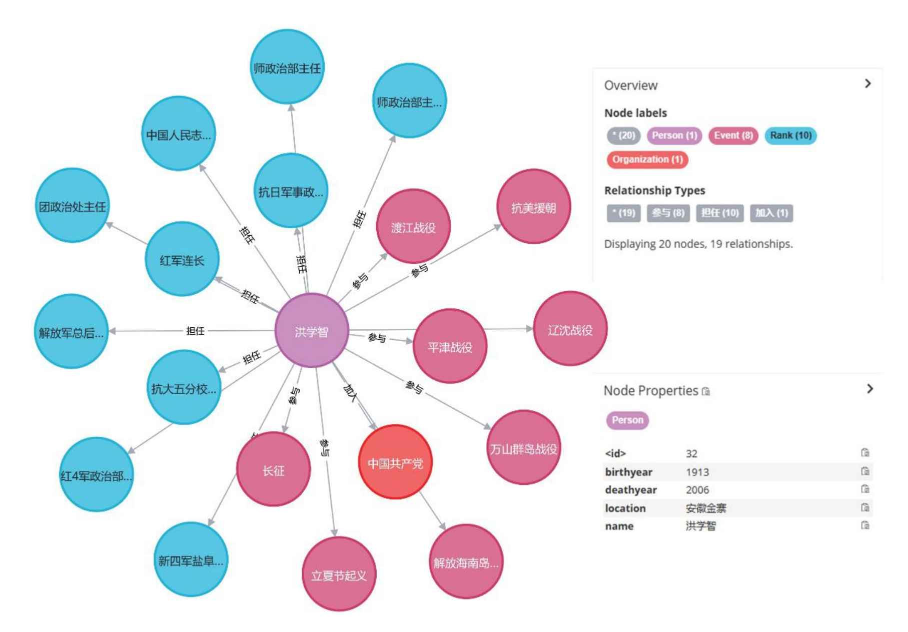


图 7 洪学智将军的履历知识图谱

图 7 左侧是以“洪学智”人物节点为中心的知识图谱，右侧为图谱节点类型与数量和“洪学智”人物节点的属性信息。由图谱可清晰获知，洪学智将军（1）基本信息：出生于 1913 年的安徽金寨，逝世于 2006 年。（2）任职经历：加入中国共产党，并担任红军连长、团政治处主任、红四军政治部主任、抗大五分校副校长、解放军总后勤部副部长等职。（3）军事经历：参与立夏节起义、红军长征、平津战役、辽沈战役、渡江战役、抗美援朝等军事战役。通过对知识图谱的网络化表示，可清晰地呈现出红色人物的基本叙事属性与履历信息，并且对红色人物信息的搜索者来说，可以快速了解红色人物的知识关联，为其寻找红色人物任职、军事经历数据提供参考。

5.2 知识图谱检索与主题关联知识发现

利用图数据库的检索与可视化功能，不仅能快速检索到红色人物的显性知识，还可以基于多主题关联检索发掘知识图谱中的隐性知识。本文择取红色人物资源叙事元素作为“原材料”，实现历史语境下皖西红色人物资源的主题检索与多元数字叙事，例如，①以人物节点为中心的知识发现，探究红色人物生平及其社会网络关系；②以事件为中心的知识发现，如探究土地革命战争、抗日战争、解放战争时期皖西红色人物群体的变化；③以机构组织为中心的知识发现，如探究中国工农红军第四军具有任职经历的红色人物，探讨其上下级与继任等关系；④以学校为中心的知识发现，如查询知识图谱中所有 “学习于”的关系，不仅可研究红色人物的受教育经历，也可以反映出各学校人才教育情况。

本研究是以人物节点为中心进行数据演示，选取革命烈士徐百川同志的履历数据为例，构建查询所用的 Cypher 检索式为“match p=(n:Person{name:'徐百川'})-[]-() return p”并展开与其相连节点的局部图谱，其中，紫色节点代表人物，桃红色节点代表事件，绿色节点代表学校，红色节点代表机构，蓝色代表职位，结果见图 8；图 9 为徐百川节点的人物属性。

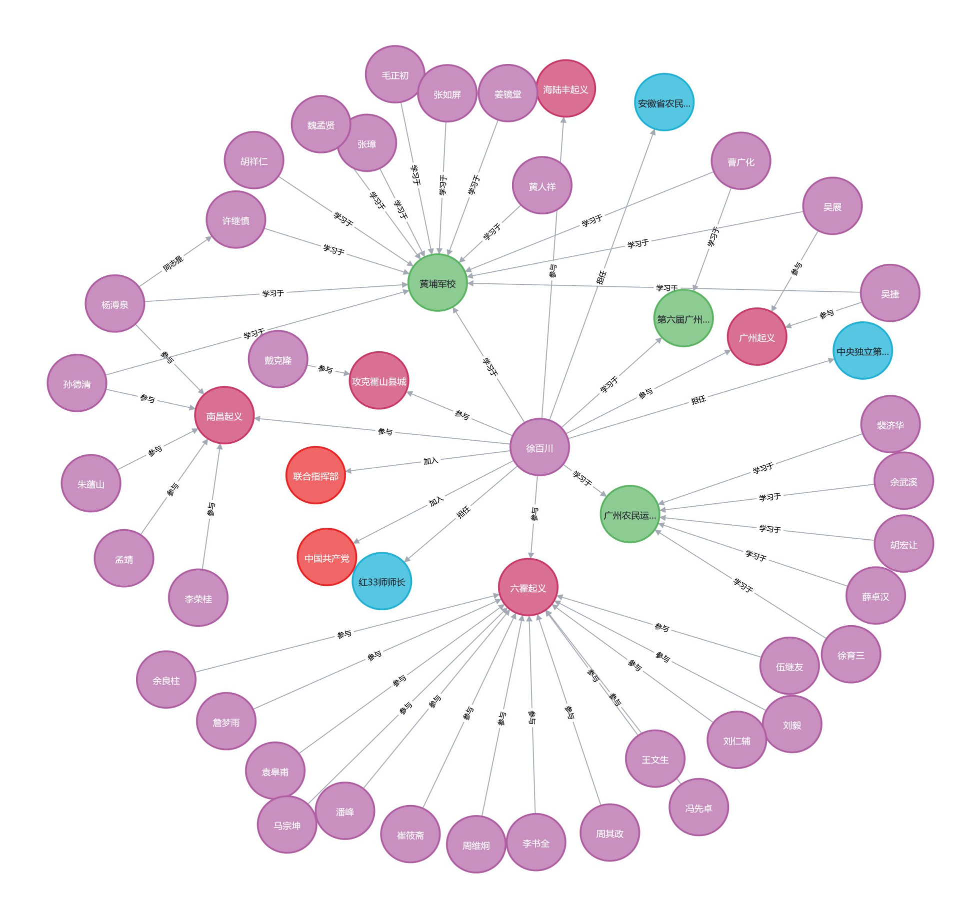


图 8 徐百川同志的履历图谱

Node Properties

Person

<id>	475
birthyear	1901
deathyear	1931
location	安徽合肥
name	徐百川
othername	张泉

图 9 节点“徐百川”的人物属性

如图 8、9 可知：第一，徐百川同志出生于 1901，逝世于 1931 年，安徽合肥人，是我党早期红色英烈；第二，曾学习于黄埔军校与第六届广州农民运动讲习所，与姜镜堂、毛正初、徐育三、余武溪等人为校友关系；第三，参与南昌起义、广州起义与六霍起义等革命战役，与朱蕴山、吴展、周维炯等人为战友关系。值得注意的是，徐百川与姜镜堂、孙德清等多人具有相似的经历，例如姜镜堂与徐百川属于同一时代，不仅都在黄埔军校深造，还同时参与了六霍起义，充分展示了历史语境下主题的内容关联，经查徐百川与姜镜堂分别担任红三十三师的师长与政治部主任，具有共事关系。

6 研究结论与启示

本研究在考虑红色人物资源来源、类型、内容及结构等多样性特征基础上，基于叙事理论构建了红色人物资源数字叙事框架，并借助“皖西红色人物”数据进行实证研究，进而利用图数据库实现了图谱可视化、知识检索与知识发现，具有理论与实践双重价值。

理论价值层面：第一，联结数字叙事与数字资源开发相关理论完善了依据红色资源“内容”和“外部”两个层次设计实体语义关联的局限，形成了可操作性的红色人物资源数字叙事框架，有助于完成数据化-情境化-故事化、数据-学术-创造的递进过程与逻辑嬗变；第二，对皖西红色人物资源进行了主题关联和语义组织等，可形成红色人物在历史时期的社会网络关系、社群动态变化、机构特征及军事教育等多种特征的静态-动态知识图谱，有助于促进历史语境下数据的解读，并辅助人文学者研究。实践价值层面：第一，较传统人工抽取与规则匹配的知识抽取方式，基于本体与 BERT 模型的三元组批量抽取模式提高了红色人物资源知识抽取效率，降低数据处理成本，促进红色文献传播与使用；第二，将静态、平面、分散的红色人物数据转变成动态、立体、聚合的知识网，提高了内容资源之间的关联程度，降低了红色人物资源检索与利用的难度，减少普通用户与交叉领域学者的研究壁垒。

参考文献

[1] 习近平.用好红色资源 赓续红色血脉 努力创造无愧于历史和人民的新业绩[J].求是，2021(19).

[2] 孙绍丹.数字人文视域下历史报纸资源语义化知识组织研究[D].长春:吉林大学,2022.

[3] 张云中,郭冬,王亚鸽,孙平.基于知识图谱的红色历史人物知识问答服务框架研究[J].图书情报工作,2021,65(16):108-117.

[4] 龙家庆,裴佳杰,张强.面向数字叙事的多模态红色馆藏资源开发策略研究[J].图书馆学研究,2023,42(5):38-46.

[5] Richardson B. Recent concepts of narrative and the narratives of narrative theory[J]. Style, 2000, 34(2): 168-175.

[6] Richardson B. Digital Fiction and the Unnatural: Transmedial Narrative Theory, Method, and Analysis[J]. Poetics Today .202243(4):773–776.

[7] 朱兰兰,段燕鸽.叙事理论在红色档案资源开发中的应用——以大别山区为例[J].档案学研究,2023,37(2):95-102.

[8] 贾君枝,崔西燕.人物本体词表之间的互操作及分类体系构建[J].情报学报,2019,38(7):731-741.

[9] 陈雅玲.基于CIDOC CRM的先秦人物知识本体构建方法研究[D].南京:南京农业大学,2019.

[10] 王楠.历史人物本体构建及其查询推理研究[D].桂林:广西师范大学,2017.

[11] Kim M, Lee D, Lee S, et al. Inference on historical factions based on multi-layered network of historical figures[J]. Expert Systems with Applications, 2020, 161: 113703.
[12] Bol P K. From Kinship to Collegiality: Changing Literati Networks, 1100-1400[J]. Journal of Historical Network Research, 2021, 5(1):87-113.
[13] 张云中,孙平.历史文化名人游学足迹知识图谱的构建与可视化[J].图书馆杂志,2021,40(09):81,87,96.
[14] 沈科杰,黄焕婷,化柏林.基于公开履历数据的人物知识图谱构建[J].数据分析与知识发现,2021,5(7):81-90.
[15] 宋雪雁,崔浩男,梁颖等.数字人文视角下名人日记资源知识发现研究——以王世杰日记为例[J].情报理论与实践,2021,44(6):105-111.
[16] 于娟,黄恒琪,席运江,朱正祥.基于图数据库的人物关系知识图谱推理方法研究[J].情报科学,2019,37(10):8-12.
[17] 赵星,李书宁,肖亚男.数字人文视域下基于多源数据融合的人物专题数据库建设——以上海图书馆 2018 开放数据应用开发竞赛作品“树人者”为例[J].图书馆杂志,2019,38(12):45-51.
[18] 夏翠娟,刘炜,陈涛,张磊.家谱关联数据服务平台的开发实践[J].中国图书馆学报,2016,42(03):27-38.
[19] Cuijuan X, Wei L, Lei Z. Implementation of a linked data-based genealogy knowledge service platform for digital humanities[J]. Data and Information Management, 2018, 2(1):15-26.
[20] 牛力,刘慧琳,高晨翔.数字记忆视角下的学术名人知识库研究[J].情报理论与实践,2022,45(9):164-172.
[21] 窦鹏.红色人物数据库建设实践与启示——以《陕甘宁边区红色记忆・人物库》为例[J].图书馆学刊,2015,37(12):36-39.
[22] 王乐.高校红色专题特藏的价值挖掘——以复旦大学图书馆为例[J].大学图书馆学报,2022,40(1):97,101,114.
[23] 刘倩倩,夏翠娟,朱武信.红色文化传承视域下的红色文献服务平台建设实践与思考[J].信息资源管理学报,2021,11(4):17,24,32,16.
[24] 孙烈涛.红色文献资源数字化建设实践探索——以梅园新村纪念馆周恩来图书馆数字化建设为例[J].新世纪图书馆,2021(04):48-52.
[25] 刘伟丽. 中共一大人物知识图谱构建研究[D].保定:河北大学,2021.
[26] 周林兴,姜璐.红色档案资源开发中的叙事表达研究[J].档案学研究,2022,36(4):4-9.
[27] 王强.讲好黄河故事之数字人文语境下黄河红色文化资源保护与育人价值释放[J].档案管理,2022,41(6):122-125.

[28] Barber J F. Digital storytelling: New opportunities for humanities scholarship and peda gogy[J]. Cogent Arts & Humanities, 2016, 3(1): 1181037.

[29] 付雅明,张永娟,刘炜等.数字叙事作为数字人文方法：现状与可能[J].图书情报工作,2022, 66(14):10-19.

[30] 张斌,李子林.图档博机构“数字叙事驱动型”馆藏利用模型[J].图书馆论坛,2021,41(5):30 -39.

[31] 徐孝娟,孙爱华,史如菊,陈涛.数字人文视角下皖西红色文献知识本体及其应用[J/OL].图书馆论坛.[2023-07-10].http://kns.cnki.net/kcms/detail/44.1306.G2.20221226.1056.002.html.

[32] 宋雪雁,张伟民,张祥青.基于档案文献的清代祭祀礼器知识图谱构建研究[J].图书情报工作,2022,66(3):140-151.

[33] 夏翠娟,刘炜,张磊等.基于书目框架(BIBFRAME)的家谱本体设计[J].图书馆论坛,2014,34 (11):5-19.

[34] XIA C,WANG L,LIU W. Shanghai memory as a digital humanities platform to rebuil d the history of the city[J].Digital scholarship in the humanities,2021,36(4):841-857.

[35] 牛力,高晨翔,张宇锋等.发现、重构与故事化:数字人文视角下档案研究的路径与方法[J].中国图书馆学报,2021,47(1):88-107.

[36] 张旭.浅析皖西红色文化资源的内涵、特征及传承开发[EB/OL].[2023-07-10].https://hswx.wxc.edu.cn/2019/0326/c5005a111030/page.htm.

[37] 黄茜茜,杨建林.基于司法判决书的知识图谱构建与知识服务应用分析[J].情报科学,2022,40(2):133-140.

[38] 宋雪雁,张伟民,张祥青.基于档案文献的清代祭祀礼器知识图谱构建研究[J].图书情报工作,2022,66(3):140-151.

作者贡献说明：徐孝娟：确定选题，论文框架设计，论文修改；孙爱华，数据收集，程序设计，论文初稿撰写及修改；陈涛，提出修改意见。

Digital Storytelling Framework for Red Character Resources and Its Empirical Research *

Xu Xiaojuan[1] Sun Aihua[1] Chen Tao[2]

1 School of Management, Anhui University，Anhui 230039

2 School of Management, Zhongshan University, Guangzhou,510006

Abstract: [Purpose/significance] Digital Storytelling as a kind of online narrative is condu cive to enhancing the interpretability and human-computer interactivity of red people's reso urces, which provides the possibility of optimizing the intellectual development of red peo ple's resources. **[Method/process]** On the basis of combing the knowledge-based developm

ent of historical character resources and the research related to red characters, the basic framework of digital storytelling of red character resources composed of narrative resource integration layer, narrative element extraction layer and narrative work generation layer is proposed, and then the knowledge mapping as a narrative tool is used as an example for empirical research on the "Red Characters of Western Anhui Province". **[Results/conclusion]** The digital storytelling framework of red people's resources can improve the current limitations of designing semantic association of entities mainly based on the "content" and "external" levels of red people's resources, and form an operable structural process, which provides reference for the development of red people's resources' knowledge; compared with the traditional manual extraction and rule-matching knowledge extraction, the ternary batch extraction mode based on ontology and BERT model improves the efficiency of red people's resources' knowledge extraction; and with the help of the knowledge mapping to achieve the red people's resources' visualization and knowledge discovery of thematic narratives, it proves the value of the application of digital storytelling in the red people's resources' knowledge development and organization.
Key words: red character resources; digital storytelling framework; knowledge mapping thematic; knowledge discovery